

# Random Popular Matchings with Incomplete Preference Lists

Suthee Ruangwises\*

Toshiya Itoh†

## Abstract

For a set  $A$  of  $n$  people and a set  $B$  of  $m$  items, with each person having a preference list that ranks a number of items in order of preference, we consider the problem of matching every person with a unique item. A matching  $M$  is called *popular* if for any other matching  $M'$ , the number of people who prefer  $M$  to  $M'$  is not less than the number of those who prefer  $M'$  to  $M$ . For given  $n$  and  $m$ , consider the probability of existence of a popular matching when each person's preference list is independently and uniformly generated at random. Previously, Mahdian showed that in the case that people's preference lists are *strict* (containing no ties) and *complete* (containing all items in  $B$ ), if  $\alpha = m/n > \alpha_*$ , where  $\alpha_* \approx 1.42$  is the root of equation  $x^2 = e^{1/x}$ , then a popular matching exists with high probability; and if  $\alpha < \alpha_*$ , then a popular matching exists with low probability. The point  $\alpha_*$  can be regarded as a transition point, at which the probability of existence of a popular matching rises from asymptotically zero to asymptotically one. In this paper, we introduce transition points in more general cases when people's preference lists are not complete. In particular, we show that in the case that each person has a preference list of length  $k$ , if  $\alpha > \alpha_k$ , where  $\alpha_k \geq 1$  is the root of equation  $xe^{-1/2x} = 1 - (1 - e^{-1/x})^{k-1}$ , then a popular matching exists with high probability; and if  $\alpha < \alpha_k$ , then a popular matching exists with low probability.

**Keywords:** popular matching, incomplete preference lists, transition point, complex component

## 1 Introduction

Consider the problem of assigning people to positions, with each person having a preference list that ranks the positions by order of preference. This simple problem models many important real-world situations, such as the assignment of graduates to training positions [8] and families to government-subsidized housing [17].

---

\*Department of Information Processing, Tokyo Institute of Technology, Yokohama, Japan; ruangwises.s.aa@m.titech.ac.jp

†Department of Mathematical and Computing Science, Tokyo Institute of Technology, Yokohama, Japan; titoh@c.titech.ac.jp

The main target of such problems is to find the “optimal” matching in each situation. Various definitions of optimality have been proposed. The least restrictive one is *Pareto optimality* [1, 2, 16]. A matching  $M$  is Pareto optimal if there is no other matching  $M'$  such that at least one person prefers  $M'$  to  $M$  but no one prefers  $M$  to  $M'$ . Other stronger definitions include *rank-maximality* [9] (allocating maximum number of people to their first choice, then maximum number to their second choices, and so on), and *popularity* [6, 3] defined below.

## 1.1 Popular Matching

Consider a set  $A$  of  $n$  people and a set  $B$  of  $m$  items, with  $\alpha = m/n$ . Throughout this paper, we assume that  $m \geq n$  and thus  $\alpha \geq 1$ . Each person in  $A$  has a preference list that ranks a number of items in order of preference. A preference list is *strict* if it does not contain ties, and is *complete* if it contains all items in  $B$ . We want to match every person with a unique item. In a matching  $M$ , for each person  $a \in A$  and item  $b \in B$ , let  $M(a)$  denote an item matched with  $a$ , and  $M(b)$  denote a person matched with  $b$ . (For convenience, let  $M(b)$  be *null* for an unmatched item  $b$ .)

Let  $r_a(b)$  be the rank of item  $b$  in  $a$ 's preference list. (The most preferred item has rank 1, the second most preferred item has rank 2, and so on.) If  $b$  is not in  $a$ 's preference list, then  $r_a(b) = \infty$ . For any pair of matchings  $M$  and  $M'$ , we define  $\phi(M, M')$  to be the number of people who prefer  $M$  to  $M'$ , i.e.  $\phi(M, M') = |\{a \in A | r_a(M(a)) < r_a(M'(a))\}|$ . We define a matching  $M$  to win over a matching  $M'$  if there are more people who prefer  $M$  to  $M'$  than those who prefer  $M'$  to  $M$ , i.e.  $\phi(M, M') > \phi(M', M)$ . A *popular matching* is a matching that does not lose to any other matching. Note that a popular matching may or may not exist.

A probabilistic variant of this problem, the random popular matching problem, studies the probability that a popular matching exists in a random instance for each value of  $n$  and  $m$ , when each person's preference list is defined independently by selecting the first item  $b_1 \in B$  uniformly at random, the second item  $b_2 \in B \setminus \{b_1\}$  uniformly at random, the third item  $b_3 \in B \setminus \{b_1, b_2\}$  uniformly at random, and so on.

## 1.2 Related Work

The concept of popularity of a matching was first introduced by Gardenfors [6] in the context of the stable marriage problem. Abraham et al. [3] presented the first polynomial time algorithm to find a popular matching in a given instance, or to report that none exists. The algorithm runs in  $O(m + n)$  time when the preference lists contain no ties, and in  $O(m\sqrt{n})$  time when the preference lists contain ties. Later, Mestre [15] generalized the algorithm to find a popular matching in the case that people are given different weights when determining the winner of two matchings. That algorithm runs in  $O(m + n)$  time when ties are not allowed, and in  $O(\min(k\sqrt{n}, n)m)$  time when ties are allowed, with  $k$  being the number of distinct weights.

A variant of this problem, known as capacitated house allocation problem, allows an item to be matched with more than one person. Manlove and Sng [13] presented an algorithm to determine whether a popular matching exists for this setting. The algorithm runs in  $O(\sqrt{C}n_1 + m)$  time when ties are not allowed, and in  $O((\sqrt{C} + n_1)m)$  when ties are allowed, where  $C$  is the total capacity,  $n_1$  is the number of people, and  $m$  is the total length of people's preference lists. The notion of a popular matching also applies when the preference list are two-sided (matching people with people), both in the bipartite graph (marriage problem) and non-bipartite graph (roommates problem). Biró et al. [5] developed an algorithm to test popularity of a matching in these two settings and also proved that determining whether a popular matching exists in these settings is an NP-hard problem when ties are allowed.

While a popular matching does not always exist, McCutchen [14] introduced two measures of the *unpopularity* of a matching, the unpopularity factor and the unpopularity margin, and showed that the problem of finding a matching that minimizes either measure is an NP-hard problem. Huang et al. [7] later gave algorithms to find a matching with bounded values of these measures in certain instances. Kavitha et al. [11] introduced the concept of a *mixed matching*, which is a probability distribution over matchings, and proved that a mixed matching that is popular always exists.

### 1.3 Our Results

For the probabilistic variant of strict and complete preference lists, Mahdian [12] proved that if  $\alpha = m/n > \alpha_*$ , where  $\alpha_* \approx 1.42$  is the root of equation  $x^2 = e^{1/x}$ , then a popular matching exists with high probability ( $1 - o(1)$  probability) in a random instance. On the other hand, if  $\alpha < \alpha_*$ , a popular matching exists with low probability ( $o(1)$  probability). The point  $\alpha = \alpha_*$  can be regarded as a transition point, at which the probability rises from asymptotically zero to asymptotically one. Itoh and Watanabe [10] later studied the case when people are given two weights  $w_1, w_2$  with  $w_1 \geq 2w_2$ , and found the transition point around  $\alpha = \Theta(n^{1/3})$  in that case.

The case that preference lists are *incomplete*, with every person's preference list has the same length  $k$ , was mentioned by Mahdian [12] and was simulated by Abraham et al. [3], but the exact transition points had not been found yet. In this paper, we study that case and discover that the transition point occurs around  $\alpha_k$ , where  $\alpha_k \geq 1$  is the root of equation  $xe^{-1/2x} = 1 - (1 - e^{-1/x})^{k-1}$ . In particular, we prove that for  $k \geq 4$ , if  $\alpha > \alpha_k$ , then a popular matching exists with high probability; and if  $\alpha < \alpha_k$ , then a popular matching exists with low probability. For  $k \leq 3$ , in which the equation does not have a solution in  $[1, \infty)$ , a popular matching always exists with high probability for every value of  $\alpha \geq 1$ .

## 2 Preliminaries

We first consider the case that every person's preference list is strict (containing no ties) and complete (containing all items in  $B$ ).

## 2.1 A-Perfect Matching

For each person  $a \in A$ , let  $f(a)$  be the item at the top of  $a$ 's preference list. Let  $F$  be the set of an item  $b \in B$  such that there exists a person  $a' \in A$  with  $f(a') = b$ , and let  $S = B - F$ . Then, for each person  $a \in A$ , let  $s(a)$  be the highest ranked item in  $a$ 's preference list that is in  $S$ .

**Definition 1.** A matching  $M$  is *A-perfect* if every person  $a \in A$  is matched with either  $f(a)$  or  $s(a)$ .

In 2005, Abraham et al. proved the following lemma.

**Lemma 1.** [3] In a given instance with strict and complete preference lists, a popular matching exists if and only if an *A-perfect* matching exists.

## 2.2 Top-Choice Graph with Complete Preference Lists

From a given instance, we construct a *top-choice graph*, a bipartite graph with parts  $B$  and  $S$  such that each person  $a \in A$  corresponds to an edge connecting  $f(a) \in B$  and  $s(a) \in S$ . Note that multiple edges are allowed in this graph. In 2006, Mahdian proved the following lemma.

**Lemma 2.** [12] In a given instance with strict and complete preference lists, an *A-perfect* matching exists if and only if its top-choice graph does not contain a *complex component*, i.e. a connected component with more than one cycle.

We define  $G(x, y, z)$  to be a bipartite graph with  $x$  and  $y$  vertices in the first and second part, respectively, with  $z$  edges selected independently and uniformly at random from the set of all possible  $xy$  edges. Mahdian also proved the following lemmas.

**Lemma 3.** [12] Suppose that we pick  $p$  elements from the set  $\{1, \dots, q\}$  independently and uniformly at random. Let a random variable  $X$  be the number of elements in the set that are not picked. Then,  $\mathbb{E}[X] = e^{-p/q}q - \Theta(1)$  and  $\text{Var}[X] < \mathbb{E}[X]$ .

**Lemma 4.** [12] Suppose that  $\alpha = m/n$ , and  $E$  is an arbitrary event defined on graphs. If the probability of  $E$  on the random graph  $G(m, h, n)$  is at most  $O(1/n)$  for every fixed integer  $h \in [e^{-1/\alpha}m - m^{2/3}, e^{-1/\alpha}m + m^{2/3}]$ , then the probability of  $E$  on the top-choice graph  $H$  is at most  $O(n^{-1/3})$ .

**Lemma 5.** [12] Suppose that  $\alpha = m/n$ , and  $h$  is an arbitrary number in  $[e^{-1/\alpha}m - m^{2/3}, e^{-1/\alpha}m + m^{2/3}]$ . Then, the probability that  $G(m, h, n)$  contains a complex component is at most  $O(1/n)$ .

Lemmas 1-5 together can conclude the following theorem.

**Theorem 1.** [12] In a random instance with strict and complete preference lists, if  $\alpha > \alpha_*$ , where  $\alpha_* \approx 1.4215$  is the solution of the equation  $x^2 e^{-1/x} = 1$ , then a popular matching exists with high probability.

Theorem 1 serves as an upper bound of the transition point in the case of strict and complete preference lists. On the other hand, the lower bound of this case was also mentioned by Mahdian [12] along with rough ideas of the proof, but the full details of the proof was not included.

**Theorem 2.** [12] In a random instance with strict and complete preference lists, if  $\alpha < \alpha_*$ , then a popular matching exists with low probability.

### 3 Incomplete Preference Lists

The previous section shows past results in the case that preference lists are strict and complete. However, in many real-world situations, preference lists may not be complete.

In the case that the preference lists are not complete, we will consider the case that every person's preference list has equal length  $k$ .

**Definition 2.** For a positive integer  $k \leq m$ , an *instance with  $k$ -incomplete preference lists* is an instance with every person's preference list having length exactly  $k$ .

For each person  $a \in A$ , let  $P_a$  be the set of items in  $a$ 's preference list. Then, let  $A_1 = \{a \in A | P_a \subseteq F\}$  and  $A_2 = \{a \in A | P_a \not\subseteq F\}$ , and let  $\beta = \frac{|A_2|}{|A|}$ . Note that the notion  $f(a)$  introduced in previous section is still well-defined for every  $a \in A$ , but  $s(a)$  is well-defined for only  $a \in A_2$ .

We first prove the necessary and sufficient conditions that a matching  $M$  is popular in an instance with strict and  $k$ -incomplete preference lists, as stated in the following lemma.

**Lemma 6.** In an instance with strict and  $k$ -incomplete preference lists, a matching  $M$  is a popular matching if and only if

1. For each person  $a \in A_1$ ,  $a$  is matched with either  $f(a)$  or an item  $b \notin P_a$ ,
2. For each person  $a \in A_2$ ,  $a$  is matched with either  $f(a)$  or  $s(a)$ , and
3. For each item  $b \in F$ ,  $b$  is matched with  $a \in A$  such that  $f(a) = b$ .

*Proof.* Suppose that the three conditions are met for a matching  $M$ . Consider any other matching  $M'$ . Assume that there is a person  $a$  that prefers  $M'$  to  $M$ .

Case 1:  $a \in A_1$ .

Then, from the first condition and the assumption that a person  $a$  prefers  $M'$  to  $M$ , we have  $M(a) \notin P_a$  and  $M'(a) \in P_a$ . So,  $M'(a) \in F$ . Let  $a' = M(M'(a))$ . From the third condition, we have  $M(a') = f(a')$  and  $M'(a') \neq f(a')$  (because  $f(a') = M(a') = M(M(M'(a))) = M'(a)$  is already matched with  $a$  in  $M'$ ), so  $a'$  prefers  $M$  to  $M'$ .

Case 2:  $a \in A_2$ .

Then, from the second condition and the assumption that a person  $a$  prefers  $M'$  to  $M$ , we have  $M(a) = s(a)$  and  $M'(a) \in F$ . Let  $a' = M(M'(a))$ . From the third condition, we have  $M(a') = f(a')$  and  $M'(a') \neq f(a')$  (because  $f(a') = M(a') = M(M(M'(a))) = M'(a)$  is already matched with  $a$  in  $M'$ ), so  $a'$  prefers  $M$  to  $M'$ .

In both cases, for any person  $a$  that prefers  $M'$  to  $M$ , there must be another unique person  $a'$  who prefers  $M$  to  $M'$ . Therefore, we can conclude that  $M$  is a popular matching.

On the other hand, suppose that  $M$  is a popular matching.

First, consider each item  $b \in F$ . If  $f(M(b)) \neq b$ , let  $a \in A$  such that  $f(a) = b$ . We promote  $a$  to match with  $b$ , promote  $M(b)$  to match with  $f(M(b))$ , and set  $M(f(M(b)))$  (if any) to match with any available item. By doing this, we make two people more satisfied and at most one person less satisfied. So,  $M$  is not a popular matching, a contradiction. Therefore, the third condition must be held.

The first condition can be obtained directly from the third condition. (If  $a \in A_1$  is matched with  $b \in P_a$  that is not  $f(a)$ , then  $b \in F$  but  $b$  is not matched with  $a'$  such that  $f(a') = b$ , a contradiction.)

Now we will prove the second condition. Suppose that  $a \in A_2$  and  $M(a)$  is neither  $f(a)$  nor  $s(a)$ .

Case 1:  $1 < r_a(M(a)) < r_a(s(a))$ .

Then,  $M(a) \in F$ , but is not matched with  $a'$  such that  $f(a') = M(a)$ , a contradiction to the third condition.

Case 2:  $r_a(M(a)) > r_a(s(a))$ .

Then, we can promote  $a$  to match with  $s(a)$ , promote  $M(s(a))$  (if any) to match with  $f(M(s(a)))$ , and set  $M(f(M(s(a))))$  (if any) to match with any available item. By doing this, we make two people more satisfied and at most one person less satisfied. So,  $M$  is not a popular matching, a contradiction. Therefore, the second condition must be held.  $\square$

Furthermore, in a given instance, if there exists a matching  $M$  that satisfies the second condition in Lemma 6, we can construct another matching  $M'$  that satisfies all the three conditions, thus being a popular matching, as stated in the following definition and lemma.

**Definition 3.** A matching  $M$  is  $A_2$ -perfect if every person  $a \in A_2$  is matched with either  $f(a)$  or  $s(a)$ .

**Lemma 7.** In a given instance with strict and  $k$ -incomplete preference lists, a popular matching exists if and only if an  $A_2$ -perfect matching exists.

*Proof.* Since the definition of an  $A_2$ -perfect matching is the second condition in Lemma 6, every popular matching is also an  $A_2$ -perfect matching by definition.

On the other hand, suppose that there exists an  $A_2$ -perfect matching  $M$  in a given instance. We will modify  $M$  to become a new matching  $M'$  that satisfies all three conditions in Lemma 6 by the following procedure: for each item  $b \in F$  that is unmatched or is matched with  $a \in A$  such that  $f(a) \neq b$ , we simply match  $b$  with any person  $a' \in A$  such that  $f(a') = b$ . If  $b$  is previously matched with  $a \in A$  such that  $f(a) \neq b$ ,  $a$  is now left unmatched; we then simply match  $a$  with any available item.

The new matching  $M'$  clearly satisfies the third condition in Lemma 6. We will prove that the second condition still holds. First, we claim that during the procedure, in the case that  $b$  is previously matched with  $a \in A$  such that  $f(a) \neq b$ , then  $a$  must be in  $A_1$ . Assume, for the sake of contradiction, that  $a \in A - A_1 = A_2$ . Since  $M$  is an  $A_2$ -perfect matching,  $a$  is matched with either  $f(a)$  or  $s(a)$ . From the assumption that  $f(a) \neq b$ , we have  $b = s(a)$ , which contradicts to the assumption that  $b \in F$ . Therefore,  $a \in A_1$ , meaning that all people in  $A_2$  are not affected by the procedure and retain the same matching, thus the second condition in Lemma 6 still holds.

Finally, we will prove that the first condition in Lemma 6 also holds for  $M'$ . Assume, for the sake of contradiction, that there is a person  $a \in A_1$  that is matched with an item  $b \in P_a$  that is not  $f(a)$ . Since  $P_a \subseteq F$ , we have  $b \in F$ . But we have  $b \neq f(a)$ , a contradiction to the third condition. Therefore, the first condition also holds.

Since  $M'$  satisfies all three conditions in Lemma 6, we can conclude that  $M'$  is a popular matching.  $\square$

Note that Lemma 7 can be regarded as a generalization of Lemma 1. In fact, in the case that  $k = m$ , we have  $A_1 = \emptyset$  and  $A_2 = A$ , which implies that Lemma 7 in that case is equivalent to Lemma 1.

### 3.1 Top-Choice Graph with Incomplete Preference Lists

Like in the case with complete preference lists, we define the top-choice graph of an instance with strict and  $k$ -incomplete preference lists to be a bipartite with parts  $B$  and  $S$ . Since  $s(a)$  is well-defined for only  $a \in A_2$ , the top-choice graph in this case will contain only  $|A_2| = \beta n$  edges, with each one corresponding to each person in  $A_2$ . (Each person  $a \in A_2$  corresponds to an edge connecting  $f(a) \in B$  and  $s(a) \in S$ .)

From Lemma 2, with set  $A$  in the lemma being the set  $A_2$ , we get the following corollary.

**Corollary 1.** In a given instance with strict and  $k$ -incomplete preference lists, an  $A_2$ -perfect matching exists if and only if its top-choice graph does not contain a complex component.

Also note that Lemma 4 still holds when replacing  $G(m, h, n)$  with  $G(m, h, \beta n)$  since the number of edges does not affect the proof of the lemma in [12].

### 3.2 Value of $\beta$

We are interested in bounding the value of ratio  $\beta = \frac{|A_2|}{|A|}$ , which will be used in the proof of our main results.

The following lemma shows that, in a random instance with strict and  $k$ -incomplete preference lists,  $\beta$  lies around  $1 - (1 - e^{-1/\alpha})^{k-1}$  with high probability.

**Lemma 8.** In a random instance with strict and  $k$ -incomplete preference lists,

$$1 - (1 - e^{-1/\alpha})^{k-1} - c < \beta < 1 - (1 - e^{-1/\alpha})^{k-1} + c$$

with high probability for any constant  $c > 0$ .

*Proof.* Let  $c > 0$  be any constant. If  $k = 1$ , then we have  $P_a \subseteq F$  for every  $a \in A$ , which means  $\beta = 0$ . From now on, we will consider the case that  $k \geq 2$ .

From Lemma 3, let  $q = m$  and  $p = n$ . Then, we have

$$\begin{aligned} \mathbb{E}[|F|] &= m - \mathbb{E}[|S|] = (1 - e^{-1/\alpha})m + \Theta(1); \\ \text{Var}(|F|) &= \text{Var}(|S|) < \mathbb{E}[|S|] < c_1 \mathbb{E}[|F|], \end{aligned}$$

for some constant  $c_1 > 0$ . Let  $c' = \frac{c}{(k-1)(c+1)}$ . We claim that

$$(1 - e^{-1/\alpha})^{k-1} - c < (1 - e^{-1/\alpha} \pm c')^{k-1} < (1 - e^{-1/\alpha})^{k-1} + c,$$

where the full explanation is given in Appendix A. From Chebyshev's inequality we have

$$\Pr[||F| - \mathbb{E}[|F|]| \geq c' \cdot \mathbb{E}[|F|]] \leq \frac{\text{Var}[|F|]}{(c' \cdot \mathbb{E}[|F|])^2} \leq \frac{c_1}{c'^2 \cdot \mathbb{E}[|F|]} = O(1/n).$$

From the fact that  $\mathbb{E}[|F|] = (1 - e^{-1/\alpha})m + \Theta(1)$ , we have  $1 - e^{-1/\alpha} - c' < \frac{|F|}{m} < 1 - e^{-1/\alpha} + c'$  with high probability for sufficiently large  $m$ .

For each  $a \in A$ , and each integer  $t \geq 1$ , consider  $\Pr[a \in A_1 | |F| = t]$ . Since  $a \in A_1$  is equivalent to  $P_a \subseteq F$ , observe that the first item in  $P_a$  is always in  $F$ , and the rest of items in  $P_a$  are uniformly selected at random from the rest  $m - 1$  items in  $B$ . Thus, we have

$$\Pr[a \in A_1 | |F| = t] = \frac{(k-1)! \binom{t-1}{k-1}}{(k-1)! \binom{m-1}{k-1}} = \frac{\binom{t-1}{k-1}}{\binom{m-1}{k-1}}.$$

Since  $\binom{t-1}{k-1} / \binom{m-1}{k-1}$  converges to  $(\frac{t}{m})^{k-1}$  for very large  $m$ , it is sufficient to consider  $\Pr[a \in A_1 | |F| = t] = (\frac{t}{m})^{k-1}$ . Using this, we can prove that

$$(1 - e^{-1/\alpha})^{k-1} - c < \Pr[a \in A_1] < (1 - e^{-1/\alpha})^{k-1} + c$$



with high probability (see Appendix B for full details). This implies

$$1 - (1 - e^{-1/\alpha})^{k-1} - c < \Pr[a \in A_2] < 1 - (1 - e^{-1/\alpha})^{k-1} + c$$

with high probability. Therefore, we can conclude that

$$1 - (1 - e^{-1/\alpha})^{k-1} - c < \beta < 1 - (1 - e^{-1/\alpha})^{k-1} + c,$$

with high probability for any constant  $c > 0$ .  $\square$

## 4 Main Results

For each value of  $k$ , we want to find a transition point  $\alpha_k$  such that if  $\alpha > \alpha_k$ , then a popular matching exists with high probability; and if  $\alpha < \alpha_k$ , then a popular matching exists with low probability. The following lemmas give an upper bound and a lower bound for  $\alpha_k$ .

### 4.1 Upper Bound

**Lemma 9.** In a random instance with strict and  $k$ -incomplete preference lists, if  $\alpha e^{-1/2\alpha} > 1 - (1 - e^{-1/\alpha})^{k-1}$ , then an  $A_2$ -perfect matching exists with high probability.

*Proof.* Since  $\alpha e^{-1/2\alpha} > 1 - (1 - e^{-1/\alpha})^{k-1}$ , we can select a small enough  $\delta > 0$  such that  $\alpha e^{-1/2\alpha} > 1 - (1 - e^{-1/\alpha})^{k-1} + \delta$ . Let  $J = [1 - (1 - e^{-1/\alpha})^{k-1} - \delta, 1 - (1 - e^{-1/\alpha})^{k-1} + \delta]$ . From Lemma 8, we have  $\beta \in J$  with high probability. We claim that it is sufficient to consider  $\beta$  to be any fixed real number in  $J$  (see Appendix C for full explanation), which implies  $\alpha e^{-1/2\alpha} > \beta$ .

We construct the top-choice graph in a way similar to Mahdian's proof of Lemma 5 in [12], but our graph has  $\beta n$  edges instead of  $n$ . Let  $X$  and  $Y$  be subsets of vertices of  $G(m, h, \beta n)$  in the first and second part, respectively. Define  $BAD_{X,Y}$  to be an event that  $X \cup Y$  contains either two vertices joined by three disjoint paths or two disjoint cycles joined by a path as a spanning subgraph. We call such subgraphs *bad* subgraphs. Note that every graph that contains a complex component must contain at least one bad subgraph.

By the same reasoning as in the proof of Lemma 5 in [12], the probability that  $BAD_{X,Y}$  occurs is at most

$$2k^2 k_1! k_2! (k+1)! \binom{\beta n}{k+1} \left( \frac{1}{mh} \right)^{k+1} \leq 2k^2 k_1! k_2! \left( \frac{\beta n}{mh} \right)^{k+1},$$

when  $k_1 = |X|$  and  $k_2 = |Y|$ .

By union bound, the probability that at least one  $BAD_{X,Y}$  occurs is at most

$$\begin{aligned}
\Pr \left[ \bigvee_{X,Y} BAD_{X,Y} \right] &\leq \sum_{k_1, k_2} \binom{m}{k_1} \binom{h}{k_2} 2k^2 k_1! k_2! \left( \frac{\beta n}{mh} \right)^{k+1} \\
&\leq \sum_{k_1, k_2} \frac{m^{k_1}}{k_1!} \cdot \frac{h^{k_2}}{k_2!} \cdot 2k^2 k_1! k_2! \left( \frac{\beta}{\alpha h} \right)^{k+1} \\
&= \sum_{k_1, k_2} \frac{2k^2}{h} \cdot \left( \frac{\beta}{\alpha} \right)^{k+1} \left( \frac{m}{h} \right)^{k_1} \\
&\leq \sum_{k=1}^{\infty} \frac{O(k^2)}{n} \cdot \left( \frac{\beta}{\alpha} \right)^k \left( e^{-1/\alpha} - m^{-1/3} \right)^{-k/2} \\
&= \frac{O(1)}{n} \sum_{k=1}^{\infty} k^2 \left( \frac{\alpha^2}{\beta^2} \left( e^{-1/\alpha} - m^{-1/3} \right) \right)^{-k/2}.
\end{aligned}$$

By the assumption, we have  $\alpha^2 e^{-1/\alpha} > \beta^2$ , so  $\frac{\alpha^2}{\beta^2} (e^{-1/\alpha} - m^{-1/3}) > 1$  for large enough  $m$ , thus the above sum converges. Therefore, the probability that at least one  $BAD_{X,Y}$  happens is at most  $O(1/n)$ .

Note that Lemma 4 still holds when replacing  $G(m, h, n)$  with  $G(m, h, \beta n)$ . By Lemma 4, the top-choice graph contains a complex component with  $O(n^{-1/3}) = o(1)$  probability. Therefore, from Corollary 1 we can conclude that an  $A_2$ -perfect matching exists with high probability.  $\square$

## 4.2 Lower Bound

**Lemma 10.** In a random instance with strict and  $k$ -incomplete preference lists, if  $\alpha e^{-1/2\alpha} < 1 - (1 - e^{-1/\alpha})^{k-1}$ , then an  $A_2$ -perfect matching exists with low probability.

*Proof.* Like in the proof of Lemma 9, we can select a small enough  $\delta > 0$  such that  $\alpha e^{-1/2\alpha} < 1 - (1 - e^{-1/\alpha})^{k-1} - \delta$ . Let  $J = [1 - (1 - e^{-1/\alpha})^{k-1} - \delta, 1 - (1 - e^{-1/\alpha})^{k-1} + \delta]$ . From Lemma 8, we have  $\beta \in J$  with high probability. Again, we claim that it is sufficient to consider  $\beta$  to be any fixed real number in  $J$  (see Appendix D for full explanation), which implies  $\alpha e^{-1/2\alpha} < \beta$ .

We use the same notions as in the proof of Lemma 9, but now we are interested in an event that  $G(m, h, \beta n)$  does not contain a complex component, which we want to prove that it occurs with low probability.

We use a method similar to Galton-Watson branching process in [4, pp.183-184]. The Galton-Watson branching process is a process that generates a random graph in a breadth-first search tree manner when given an expected number of total edges. The process starts

when the starting node spawns a number of children that are ordered in some way. Then, each of its children, in that order, also spawns children by the same manner. The number of children each node spawns follows a Poisson distribution with mean  $c$ . The process may stop at some point or continues indefinitely.

For a small enough  $\epsilon > 0$ , we still have  $\alpha e^{-1/2\alpha} < (1 - \epsilon)\beta$ . Consider the construction of a bipartite graph  $G(m, h, (1 - \epsilon)\beta n)$  using Galton-Watson branching process. The number of children each node spawns follows a binomial distribution such that the expected total number of edges are  $(1 - \epsilon)\beta n$ , which becomes Poisson distribution when  $n$  is very large. Precisely, the Poisson distribution of a node in the first part has mean  $c_1 = \frac{\beta n}{m} = \frac{\beta}{\alpha}$ , and that of a node in the second part has mean  $c_2 = \frac{\beta n}{h} = \frac{\beta}{\alpha e^{-1/\alpha}}$ .

Let  $T$  be the size of the process ( $T = \infty$  if the process continues forever). Let  $z_1$  and  $z_2$  be the probability that  $T < \infty$  when starting the process at a node in the first and second part of the graph, respectively. Also, let  $Z_1$  and  $Z_2$  be the number of children the root has when starting the process at a node in the first and second part of the graph, respectively.

Given that the root has  $i$  children, in order for the branching process to be finite, all of the  $i$  branches must be finite, so we get the equations.

$$\begin{aligned} z_1 &= \sum_{i=0}^{\infty} \Pr[Z_1 = i] z_2^i \\ z_2 &= \sum_{i=0}^{\infty} \Pr[Z_2 = i] z_1^i \end{aligned}$$

Therefore,

$$z_1 = \sum_{i=0}^{\infty} \frac{c_1^i e^{-c_1}}{i!} \left( \sum_{j=0}^{\infty} \frac{c_2^j e^{-c_2} z_1^j}{j!} \right)^i = \sum_{i=0}^{\infty} \frac{c_1^i e^{-c_1}}{i!} e^{c_2(z_1-1)i} = e^{c_1(e^{c_2(z_1-1)}-1)}.$$

Setting  $y = 1 - z_1$  yields the equation

$$1 - y = e^{c_1(e^{-c_2 y} - 1)}. \quad (1)$$

Define  $g(y) = 1 - y - e^{c_1(e^{-c_2 y} - 1)}$ . We have  $g(0) = 1 - 0 - 1 = 0$  and  $g(1) < 0$ . Also, since  $c_1 c_2 = \frac{\beta^2}{\alpha^2 e^{-1/\alpha}} > 1$  by assumption, we have  $g'(0) = c_1 c_2 - 1 > 0$ . So, there must be  $y \in (0, 1)$  such that  $g(y) = 0$ , thus being a solution of (1).

So,  $\Pr[T = \infty] = y \in (0, 1)$ , when  $y$  is a solution of (1), meaning that the expected size of the largest component in the graph is at least  $ym'$ , when  $m' = m + h$  is the number of vertices in the graph. Thus, the size of the largest component is proportional to  $m$ .

Finally, consider putting the remaining  $\epsilon\beta n$  edges into the graph independently and uniformly at random. Note that if at least two of those edges land in the largest component  $C$ ,

a complex component will be created. Since  $C$  has size proportional to  $m$ , each edge has a constant probability to land in  $C$ , so the probability that at most one edge will land in  $C$  is exponentially low. Therefore,  $G(m, h, \beta n)$  does not contain a complex component with probability less than  $O(1/n)$ .

By Lemma 4, the top-choice graph does not contain a complex component with  $O(n^{-1/3}) = o(1)$  probability. Therefore, from Corollary 1 we can conclude that an  $A_2$ -perfect matching exists with low probability.  $\square$

**Note:** Note that Mahdian mentioned but did not show full details of his proof for the lower bound in [12], but it is likely that his proof uses the same technique. In the case with complete preference lists with  $\alpha e^{-1/2\alpha} < 1$ , we have  $c_1 = \frac{1}{\alpha}$  and  $c_2 = \frac{1}{\alpha e^{-1/\alpha}}$ , which we still get  $c_1 c_2 = \frac{1}{\alpha^2 e^{-1/\alpha}} > 1$ . This is a sufficient condition to reach the same conclusion that  $g'(0) > 0$ , and thus  $\Pr[T = \infty] \in (0, 1)$ .

### 4.3 Transition Point

Finally, from Lemmas 7, 9, and 10, we can conclude the following theorem.

**Theorem 3.** In a random instance with  $k$ -incomplete preference lists, if  $\alpha e^{-1/2\alpha} > 1 - (1 - e^{-1/\alpha})^{k-1}$ , then a popular matching exists with high probability. On the other hand, if  $\alpha e^{-1/2\alpha} < 1 - (1 - e^{-1/\alpha})^{k-1}$ , then a popular matching exists with low probability.

Since  $f(x) = xe^{-1/2x} - (1 - (1 - e^{-1/x})^{k-1})$  is an increasing function in  $[1, \infty)$  for every  $k \geq 1$ ,  $f(x) = 0$  can have at most one root in  $[1, \infty)$ . That root, if exists, will serve as a transition point  $\alpha_k$ . In fact, for  $k \geq 4$ ,  $f(x) = 0$  has a unique solution in  $[1, \infty)$ ; for  $k \leq 3$ ,  $f(x) = 0$  has no solution in  $[1, \infty)$ , and we have  $\alpha e^{-1/2\alpha} > 1 - (1 - e^{-1/\alpha})^{k-1}$  for every  $\alpha \geq 1$ , meaning that a popular matching always exists with high probability regardless of value of  $\alpha$ , so there is no transition point.

**Corollary 2.** In a random instance with  $k$ -incomplete preference lists with  $k \geq 4$ , if  $\alpha > \alpha_k$ , where  $\alpha_k \geq 1$  is the root of equation  $xe^{-1/2x} = 1 - (1 - e^{-1/x})^{k-1}$ , then a popular matching exists with high probability; and if  $\alpha < \alpha_k$ , then a popular matching exists with low probability. For  $k \leq 3$ , a popular matching always exists with high probability in a random instance with  $k$ -incomplete preference lists for every  $\alpha \geq 1$ .

### 4.4 Discussion

For each value of  $k \geq 4$ , the transition point occurs at the root  $\alpha_k \geq 1$  of equation  $xe^{-1/2x} = 1 - (1 - e^{-1/x})^{k-1}$  as shown in the following figure and table. Note that as  $k$  increases, the right-hand side term of the equation converges to 1, thus the transition point  $\alpha_k$  converges to Mahdian's value of  $\alpha_* \approx 1.4215$  in the case with complete preference lists.

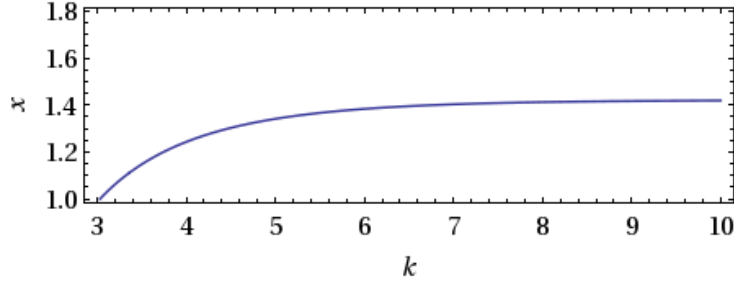


Figure 1: Plot of the equation  $xe^{-1/2x} = 1 - (1 - e^{-1/x})^{k-1}$ , showing the solution in  $[1, \infty)$  for each value of  $k$

$k$	4	5	6	7	8	9	10	...	20	...
$\alpha_k$	1.2428	1.3411	1.3835	1.4031	1.4124	1.4170	1.4193	...	1.4215	...

Table 1: Solution in  $[1, \infty)$  of the equation  $xe^{-1/2x} = 1 - (1 - e^{-1/x})^{k-1}$  for each integer value of  $k \geq 4$

## 5 Future Work

In many real-world situations, ties can and are likely to occur among each person's preference list. The random popular matching problem in the case with ties allowed was mentioned by Mahdian [12] and was simulated by Abraham et al. [3] using a parameter  $t$  to denote the probability that each entry in a preference list is tied with previous entry. Intuitively, and also confirmed by the experimental results of [3], when ties are very likely to occur ( $t$  is very close to 1), a popular matching is likely to exist even when  $\alpha = 1$ . However, the transition point for each value of  $t$ , or whether it exists at all, has still not been found yet. A possible future work is to prove the existence and find the transition point for each value of  $t$ , both in the cases with complete and incomplete preference lists.

## References

- [1] A. Abdulkadiroğlu and T. Sönmez. Random serial dictatorship and the core from random endowments in house allocation problems. *Econometrica* 66(3):689-701 (1998).
- [2] D.J. Abraham, K. Cechlárová, D.F. Manlove, and K. Mehlhorn. Pareto-optimality in house allocation problems. In *Proceedings of 15th Annual International Symposium on Algorithms and Computation (SODA)*, pages 3-15 (2004).
- [3] D.J. Abraham, R.W. Irving, T. Kavitha, and K. Mehlhorn. Popular matchings. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 424-432 (2005).
- [4] N. Alon and J. Spencer. The probabilistic method. John Wiley & Sons (2000).

- [5] P. Biró, R.W. Irving, and D. Manlove. Popular Matchings in the Marriage and Roommates Problems. In *Proceedings of the 7th International Conference on Algorithms and Complexity (CIAC)*, pages 97-108 (2010).
- [6] P. Gärdenfors. Match making: assignments based on bilateral preferences. *Behav. Sci.* 20:166-173 (1975).
- [7] C.-C. Haung, T. Kavitha, D. Michail, and M. Nasre. Bounded unpopularity matchings. In *Proceedings of the 11th Scandinavian Workshop on Algorithm Theory (SWAT)*, pages 127-137 (2008).
- [8] A. Hylland, and R. Zeckhauser. The efficient allocation of individuals to positions. *J. Polit. Econ.* 87(22):293-314 (1979).
- [9] R.W. Irving, T. Kavitha, K. Mehlhorn, D. Michail, and K. Paluch. Rank-maximal matchings. *ACM Trans. Algorithms* 2(4):602-610 (2006).
- [10] T. Itoh and O. Watanabe. Weighted random popular matchings. *Random Struct. Algorithms*, 37(4):477-494 (2010).
- [11] T. Kavitha, J. Mestre, and M. Nasre. Popular mixed matchings. In *Proceedings of the 36th International Colloquium on Automata, Languages and Programming (ICALP)*, pages 574-584 (2009).
- [12] M. Mahdian. Random popular matchings. In *Proceedings of the 7th ACM Conference on Electronic Commerce (EC)*, pages 238-242 (2006).
- [13] D. Manlove and C.T.S. Sng. Popular matchings in the weighted capacitated house allocation problem. *J. Discrete Algorithms* 8(2):102-116 (2010).
- [14] R.M. McCutchen. The least-unpopularity-factor and least-unpopularity-margin criteria for matching problems with one-sided preferences. In *Proceedings of the 15th Latin American Symposium on Theoretical Informatics (LATIN)*, pages 593-604 (2008).
- [15] J. Mestre. Weighted popular matchings. In *Proceedings of the 16th International Colloquium on Automata, Languages, and Programming (ICALP)*. pages 715-726 (2006).
- [16] A.E. Roth and A. Postlewaite. Weak versus strong domination in a market with indivisible goods. *J. Math. Econ.* 4:131-137 (1977).
- [17] Y. Yuan. Residence exchange wanted: a stable residence exchange problem. *Eur. J. Oper. Res.* 90:536-546 (1996).

## A Upper and Lower Bounds for $(1 - e^{-1/\alpha} \pm c')^{k-1}$

For  $k \geq 2$ , let  $p = 1 - e^{-1/\alpha}$  and  $c' = \frac{c}{(k-1)(c+1)}$ . We have  $0 < p < 1$  and  $0 < c' < 1$ . So,

$$\begin{aligned}
(p + c')^{k-1} &= p^{k-1} + \binom{k-1}{1} p^{k-2} c' + \binom{k-1}{2} p^{k-3} c'^2 + \dots + \binom{k-1}{k-1} c'^{k-1} \\
&\leq p^{k-1} + (k-1) p^{k-2} c' + (k-1)^2 p^{k-3} c'^2 + \dots + (k-1)^{k-1} c'^{k-1} \\
&\leq p^{k-1} + (k-1) c' + (k-1)^2 c'^2 + \dots + (k-1)^{k-1} c'^{k-1} \\
&= p^{k-1} + \frac{c}{c+1} + \left( \frac{c}{c+1} \right)^2 + \dots + \left( \frac{c}{c+1} \right)^{k-1} \\
&\leq p^{k-1} + \frac{c}{c+1} + \left( \frac{c}{c+1} \right)^2 + \dots \\
&= p^{k-1} + \frac{\frac{c}{c+1}}{1 - \frac{c}{c+1}} \\
&= p^{k-1} + c.
\end{aligned}$$

Also, we have

$$\begin{aligned}
(p - c')^{k-1} &= p^{k-1} - \binom{k-1}{1} p^{k-2} c' + \binom{k-1}{2} p^{k-3} c'^2 - \dots + (-1)^{k-1} \binom{k-1}{k-1} c'^{k-1} \\
&\geq p^{k-1} - [(k-1) p^{k-2} c' + (k-1)^2 p^{k-3} c'^2 + \dots + (k-1)^{k-1} c'^{k-1}] \\
&\geq p^{k-1} - [(k-1) c' + (k-1)^2 c'^2 + \dots + (k-1)^{k-1} c'^{k-1}] \\
&= p^{k-1} - \left[ \frac{c}{c+1} + \left( \frac{c}{c+1} \right)^2 + \dots + \left( \frac{c}{c+1} \right)^{k-1} \right] \\
&\geq p^{k-1} - \left[ \frac{c}{c+1} + \left( \frac{c}{c+1} \right)^2 + \dots \right] \\
&= p^{k-1} - \frac{\frac{c}{c+1}}{1 - \frac{c}{c+1}} \\
&= p^{k-1} - c.
\end{aligned}$$

Therefore  $(1 - e^{-1/\alpha})^{k-1} - c < (1 - e^{-1/\alpha} \pm c')^{k-1} < (1 - e^{-1/\alpha})^{k-1} + c$ .

## B Upper and Lower Bounds for $\Pr[a \in A_1]$

Let  $I = [m(1 - e^{-1/\alpha} - c'), m(1 - e^{-1/\alpha} + c')]$ . We have  $|F| \in I$  with high probability.

Consider  $\Pr[a \in A_1]$ . We have

$$\begin{aligned}\Pr[a \in A_1] &= \sum_t \Pr[|F| = t] \cdot \Pr[a \in A_1 | |F| = t] \\ &= \sum_{t \in I} \Pr[|F| = t] \cdot \Pr[a \in A_1 | |F| = t] + \sum_{t \notin I} \Pr[|F| = t] \cdot \Pr[a \in A_1 | |F| = t].\end{aligned}$$

For the lower bound of  $\Pr[a \in A_1]$ , we have

$$\begin{aligned}\Pr[a \in A_1] &\geq \sum_{t \in I} \Pr[|F| = t] \cdot \Pr[a \in A_1 | |F| = t] \\ &= \sum_{t \in I} \Pr[|F| = t] \cdot \left(\frac{t}{m}\right)^{k-1} \\ &\geq \sum_{t \in I} \Pr[|F| = t] \cdot (1 - e^{-1/\alpha} - c')^{k-1} \\ &= \Pr[|F| \in I] \cdot (1 - e^{-1/\alpha} - c')^{k-1} \\ &\geq (1 - o(1))((1 - e^{-1/\alpha})^{k-1} - c)\end{aligned}$$

On the other hand, for the upper bound of  $\Pr[a \in A_1]$ , we have

$$\begin{aligned}\Pr[a \in A_1] &\leq \sum_{t \in I} \Pr[|F| = t] \cdot \Pr[a \in A_1 | |F| = t] + \sum_{t \notin I} \Pr[|F| = t] \\ &= \sum_{t \in I} \Pr[|F| = t] \cdot \left(\frac{t}{m}\right)^{k-1} + o(1) \\ &\leq \sum_{t \in I} \Pr[|F| = t] \cdot (1 - e^{-1/\alpha} + c')^{k-1} + o(1) \\ &= \Pr[|F| \in I] \cdot (1 - e^{-1/\alpha} + c')^{k-1} + o(1) \\ &\leq (1 - o(1))((1 - e^{-1/\alpha})^{k-1} + c) + o(1)\end{aligned}$$

Therefore, we can conclude that  $(1 - e^{-1/\alpha})^{k-1} - c < \Pr[a \in A_1] < (1 - e^{-1/\alpha})^{k-1} + c$  with high probability.

## C Full Explanation for the Proof of Lemma 9

For given  $n$ ,  $m$ , and  $k$ , let  $E_1$  be an event that an  $A_2$ -perfect matching exists in a random instance with strict and  $k$ -incomplete preference lists. In Lemma 9, we have proved that if  $\alpha e^{-1/2\alpha} > 1 - (1 - e^{-1/\alpha})^{k-1}$ , then  $\Pr[E_1 | \beta = t] = 1 - o(1)$  for every fixed  $t \in J$ . So, we



have

$$\begin{aligned}
\Pr[E_1] &= \sum_t \Pr[\beta = t] \cdot \Pr[E_1 | \beta = t] \\
&= \sum_{t \in J} \Pr[\beta = t] \cdot \Pr[E_1 | \beta = t] + \sum_{t \notin J} \Pr[\beta = t] \cdot \Pr[E_1 | \beta = t] \\
&\geq \sum_{t \in J} \Pr[\beta = t] \cdot \Pr[E_1 | \beta = t] \\
&\geq \Pr[\beta \in J] \cdot (1 - o(1)) \\
&= (1 - o(1))(1 - o(1)) \\
&= 1 - o(1).
\end{aligned}$$

Therefore, an  $A_2$ -perfect matching exists with high probability in a random instance.

## D Full Explanation for the Proof of Lemma 10

Like in Appendix C, for given  $n$ ,  $m$ , and  $k$ , let  $E_1$  be an event that an  $A_2$ -perfect matching exists in a random instance with strict and  $k$ -incomplete preference lists. In Lemma 10, we have proved that if  $\alpha e^{-1/2\alpha} < 1 - (1 - e^{-1/\alpha})^{k-1}$ , then  $\Pr[E_1 | \beta = t] = o(1)$  for every fixed  $t \in J$ . So, we have

$$\begin{aligned}
\Pr[E_1] &= \sum_t \Pr[\beta = t] \cdot \Pr[E_1 | \beta = t] \\
&= \sum_{t \in J} \Pr[\beta = t] \cdot \Pr[E_1 | \beta = t] + \sum_{t \notin J} \Pr[\beta = t] \cdot \Pr[E_1 | \beta = t] \\
&\leq \sum_{t \in J} \Pr[\beta = t] \cdot \Pr[E_1 | \beta = t] + \sum_{t \notin J} \Pr[\beta = t] \\
&\leq \Pr[\beta \in J] \cdot o(1) + \Pr[\beta \notin J] \\
&\leq o(1) + \Pr[\beta \notin J] \\
&= o(1) + o(1) \\
&= o(1).
\end{aligned}$$

Therefore, an  $A_2$ -perfect matching exists with low probability in a random instance.